

# Cancer data API

Using the languages, libraries, and frameworks of your choice, build an application that serves a web API.

## Submission

You may submit the application as a private GitHub repository.

## Technical requirements

- API served over a network (HTTP, REST, SOAP, gRPC, etc).
- Application runs inside a Docker container.

## API specification

The design of the API is up to you. It must perform at least the following behaviors:

- return whether a given gene is a TCGA hotspot for a given cell line
- return the list of TCGA hotspot genes for a given cell line
- return the list of cell lines that a given gene is a TCGA hotspot for

## Data

Use the cellular model mutations dataset ( `CCLC_mutations.csv` ) from the [Cancer Dependency Map \(DepMap\)](#). The full dataset can be downloaded directly [here](#). Feel free to truncate the file to the first 10,000 rows or so to make it more manageable.

## Data dictionary

The file is a matrix (gene × cell line) encoded as a CSV. Important columns include:

- `Entrez_Gene_Id` : [Entrez IDs](#) (e.g. [7105](#)) identify genes;
- `DepMap_ID` : DepMap coded identifiers for [cell lines](#) (e.g. [ACH-001113](#) );
- `isTCGAhotspot` : whether this gene is frequently mutated in this cancer cell line;

You may ignore the rest.

## Notes

There may be multiple rows for a single gene/cell line pair. If *any* of those rows has `isTCGAhotspot=true` , consider the gene a TCGA hotspot for that cell line.